Characterisation of thiamine diphosphate-dependent 4-hydroxybenzoylformate decarboxylase enzymes from *Rhodococcus jostii* RHA1 and *Pseudomonas fluorescens* Pf-5 involved in degradation of aryl-C_2 lignin degradation fragments

Zhen Wei[1], Rachael C. Wilkinson[2], Goran M. M. Rashid[1], David Brown[1], Vilmos Fülöp[2] and Timothy D.H. Bugg[1]*

1. Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK
2. School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

Corresponding author: Prof. Timothy D.H. Bugg, Department of Chemistry, University of Warwick, Coventry CV4 7AL, United Kingdom. Email T.D.Bugg@warwick.ac.uk, tel 44-2476-573018.

Running title: 4-Hydroxybenzoylformate decarboxylase

Abstract:

A thiamine diphosphate-dependent enzyme annotated as a benzoylformate decarboxylase is encoded in gene cluster ro02984-ro02986 in *Rhodococcus jostii* RHA1 previously shown to generate vanillin and 4-hydroxybenzaldehyde from lignin oxidation, and a closely related gene cluster is also found in the genome of *Pseudomonas fluorescens* Pf-5. Two hypotheses for possible pathways involving a thiamine diphosphate-dependent cleavage, either C-C cleavage of a ketol or diketone aryl C_3 substrate, or decarboxylation of an aryl C_2 substrate, were investigated by expression and purification of the recombinant enzymes, and expression of dehydrogenase and oxidase enzymes also found in the gene clusters. The ThDP-dependent enzymes showed no activity for cleavage of aryl C_3 ketol or diketone substrates, but showed activity for decarboxylation of benzoylformate and 4-hydroxybenzoylformate. A flavin-dependent oxidase encoded by gene ro02984 was found to oxidise either mandelic acid or phenylglyoxal. The crystal structure of the *P. fluorescens* decarboxylase enzyme was determined at 1.69 Å resolution, showing similarity to known benzoylformate decarboxylase enzymes. The *P. fluorescens* decarboxylase enzyme showed enhanced carboligase activity between vanillin and acetaldehyde, rationalised by the presence of alanine vs serine at residue 73 in the enzyme active site, which was investigated further by site-directed mutagenesis of this residue. A hypothesis for a pathway for degradation of aryl-C_2 fragments arising from oxidative cleavage of phenylcoumaran and diarylpropane structures in lignin is proposed.

Keywords: benzoylformate decarboxylase; thaimine diphosphate; *Rhodococcus jostii* RHA1; *Pseudomonas fluorescens* Pf-5; lignin degradation.

Database: structural data are available in the PDB under the accession number 6QSI

Introduction

The aromatic heteropolymer lignin comprises 15-30% of plant cell wall lignocellulose found in all plant biomass. There is current interest in the use of biotechnology to convert lignin into renewable chemicals, as part of a lignin-based biorefinery to produce renewable fuels and chemicals from plant biomass (*1*). However, the microbial degradation of lignin in soil is still not fully understood at the molecular level. In basidiomycete fungi, extracellular lignin peroxidases, manganese peroxidases and laccases for lignin oxidation have been well characterised (*2,3*). Since 2011, bacterial dye-decolorizing peroxidases have been discovered in *Rhodococcus jostii* RHA1 (*4*), *Amycolatopsis* sp. 75iv2 (*5*) and *Pseudomonas fluorescens* Pf-5 (*6*) that can oxidise the lignin polymer, and multi-copper oxidases have been identified in *Streptomyces coelicolor* (*7*) and *Ochrobactrum* sp (*8*) that can oxidise lignin. It is likely that the lignin polymer is oxidised by these enzymes extracellularly, to generate oxidised lignin fragments that are then taken up into the cell, and metabolised to monocyclic aromatic compounds. The pathways for bacterial degradation of aromatic compounds are well understood (*9*), but the pathways by which oxidised lignin fragments are converted into aromatic monomers are poorly understood (*10*).

Generation of vanillic acid as a metabolite of lignin oxidation has been commonly observed (*11*), which is converted via demethylation into protocatechuic acid (*12*). Gene deletion of gene ro02986 encoding vanillin dehydrogenase in *R. jostii* RHA1 gives a mutant strain that accumulates vanillin and 4-hydroxybenzaldehyde when grown in minimal media containing wheat straw lignocellulose (*13*), demonstrating that one avenue for lignin breakdown in bacteria proceeds through vanillin and vanillic acid. Vanillin can be formed via Cα-Cβ oxidative cleavage of lignin model compounds by fungal lignin peroxidases (*2*) and bacterial dye-decolorizing peroxidases (*4*), but there may be other lignin degradation pathways that also converge on vanillin and vanillic acid. A thorough understanding of pathways for lignin oxidation is vitally important for metabolic engineering of lignin degradation into high value chemicals (*1*).

The gene cluster in *R. jostii* RHA1 containing the ro02986 vanillin dehydrogenase gene contains several other genes whose function may be linked, suggesting that this cluster may encode such a breakdown pathway. These include gene ro02985 annotated as a thiamine diphosphate-dependent benzoylformate decarboxylase, and gene ro02984 annotated as a glycolate oxidase, shown in Figure 1. The enzyme benzoylformate

decarboxylase catalyses the decarboxylation of benzoylformate to benzaldehyde, and is used in *Pseudomonas putida* for the degradation of mandelic acid (*14*), however, the *R. jostii* gene cluster is lacking genes for mandelic acid racemisation and oxidation, suggesting that the function of this cluster may be different to that in *P. putida*. The presence of a gene for glycolate oxidation to oxalic acid is consistent with the observation of oxalic acid as a metabolite in lignin-degrading bacteria (*15*), and suggests the generation of a two-carbon fragment. A related cluster was also found in the genome of *Pseudomonas fluorescens* Pf-5, a bacterium from which we have previously isolated a lignin-oxidising DyP peroxidase enzyme (*6*), as shown in Figure 1.

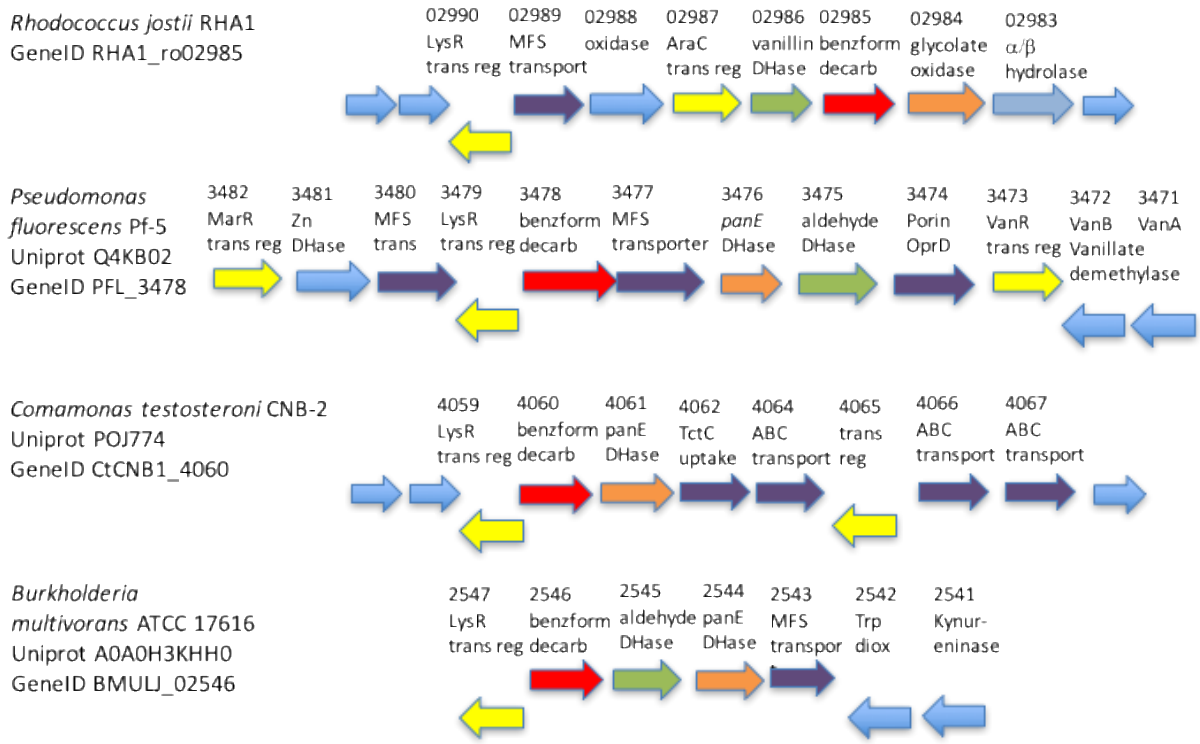


Figure 1 Organisation of gene clusters containing benzoylformate decarboxylase genes in *Rhodococcus jostii* RHA1, *Pseudomonas fluorescens* Pf-5, *Comamonas testosteroni* CNB-2, and *Burkholderia multivorans* ATCC 17616. Key: red, benzoylformate decarboxylase genes; green, aldehyde dehydrogenase genes; brown, oxidase or PanE dehydrogenase genes; yellow, transcriptional regulator genes; purple, transporter genes.

Two hypotheses were therefore devised for the function of the ThDP-dependent enzyme encoded by *R. jostii* ro02985 and *P. fluorescens* PFL_3478, as shown in Figure 2:

A) C-C cleavage of an aryl C_3 ketol or diketone intermediate, akin to a thiamine diphosphate-dependent transketolase enzyme (*16*), generating glycoaldehyde as a by-product, which is then oxidised by the putative glycolate oxidase to oxalic acid; B) decarboxylation of a substituted benzoylformate, generated via oxidation of an aryl C_2 fragment. Hypothesis A uses an oxidised aryl C_3 substrate which is consistent with the general aryl C_3 structure of lignin (see Figure 2A), and the presence of a glycolate oxidase enzyme, also a lignin oxidation product containing similar keto-diol sidechain has been observed previously from treatment of wheat straw lignocellulose by *P. fluorescens* Dyp1B (*6*). Hypothesis B is consistent with the annotation of the ThDP-dependent enzyme as a benzoylformate decarboxylase (see Figure 2), but it is less clear how an aryl C_2 fragment would be generated from lignin oxidation.

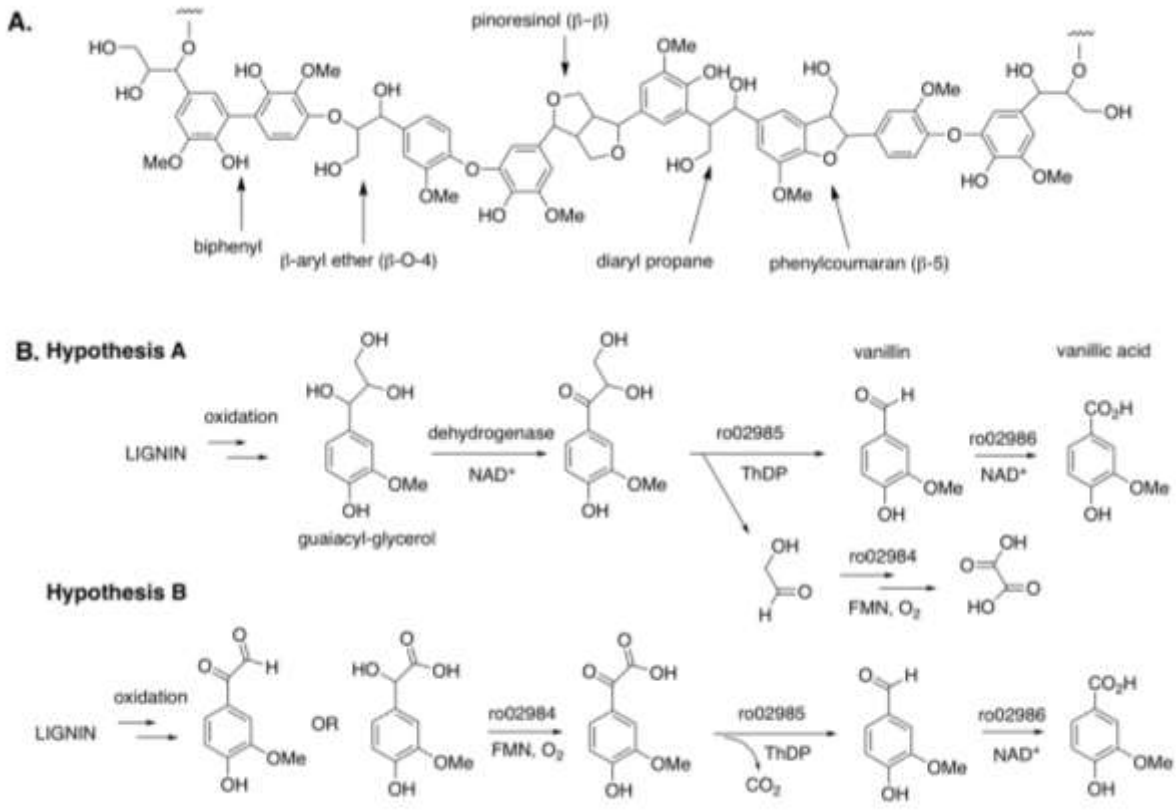


Figure 2. A. Representative structure of polymeric lignin, showing different structural units found in lignin structure. B. Hypotheses A & B for possible functions of enzymes encoded by *R. jostii* ro02984-ro02986 genes.

This paper will describe the investigation of these two hypotheses using recombinant enzymes from *R. jostii* RHA1 and *P. fluorescens* Pf-5, and structure determination of the *P. fluorescens* benzoylformate decarboxylase enzyme.

Materials and Methods

Materials. Guaiacyl-glycerol was synthesised from ferulic acid using the method of Adler & Yllner (*17*). ^{1}H NMR (300 MHz, CD_3OD) δ_H 7.03 (1H, s, aryl H-2), 6.84 (1H, dd, J = 6.8, 1.2 Hz, aryl H-6), 6.78 (1H, d, J = 6.8 Hz, aryl H-5), 4.55 (1H, d, J = 6.7 Hz, $C_\alpha\underline{H}$(OH)), 3.88 (3H, s, OCH_3), 3.5-3.8 (3H, m, $C_\beta\underline{H}$(OH) & $C_\gamma\underline{H_2}$OH) ppm. ^{1}H NMR spectrum shown in Supporting Information. 1-(4-Hydroxy-3-methoxyphenyl)propane-1,2-dione was a gift from Prof. Jeremy Luterbacher, Laboratory of Sustainable and Catalytic Processing, EPFL, Switzerland. Other substrates were purchased from Sigma-Aldrich and Chemfaces.

Gene cloning and enzyme overexpression. Genomic DNA was extracted using the Wizard_Genomic DNA Purification Kit from Promega, from a total of 1 ml of overnight culture of *R. jostii* RHA1 and *P. fluorescens* Pf-5 in Luria-Bertani broth. The amplified genes were cloned using the Champion pET151 directional TOPO Expression Kit (Invitrogen) into expression vector pET151 or Expresso T7 SUMO Cloning and Expression System Kit (Lucigen) into pETite™ N-His SUMO Vector. The designed primers for each gene are shown in Table 1. Then the constructs were transformed into *E. coli* TOP10 competent cells (Invitrogen) and extracted plasmids from the obtained colonies were sent for DNA sequencing in order to confirm the accuracy of the sequence. The recombinant plasmids were then transformed into *E. coli* BL21 (Invitrogen), for protein expression. For expression of each recombinant gene, a 15 ml starter culture was grown in Luria-Bertani broth in the presence of antibiotics 100 μg/ml ampicillin for vector pET151 and 30 μg/ml kanamycin for vector pETite™ N-His SUMO for 6 h at 37 °C, then added to 1 L Luria–Bertani broth for 4 h at 37 °C, and finally the cells were induced by adding 0.25 mM final concentration of isopropyl-β-D-thiogalactopyranoside (IPTG) and shaken overnight at 15 °C. Then the His-tagged proteins were purified to near homogeneity by using immobilized ion affinity chromatography (IMAC) under native conditions (Ni-NTA column, Qiagen), as described previously (*18*). RjBFDC and PfBFDC were purified in the presence of 0.1 mM thiamine diphosphate. Oxidase ro02984 was expressed in the presence of 0.1 mM flavin

mononucleotide (FMN). Fractions were collected and analysed by SDS-PAGE gel electrophoresis. The protein concentration was detected using Bradford protein assay kit (BioRad) and by absorbance at 280 nm using Beer's Law with theoretical mass extinction coefficient of 65235 M^{-1} cm^{-1} (ExPASY) for crystallisation.

Table 1 List of vectors and primers used for PCR gene amplification

Bacteria	gene	Enzyme	vector	Forward (5'-3')	Reverse (5'-3')
R. jostii RHA1	ro02984	Oxidase	pETSUMO	CGCGAACAGATTGGAGGTAATCTCAT CGAGACACTGAAG	GTGGCGGCCGCTCTATTAGAACCC ACGGCCGGGATTGAG
	ro02985	BFDC	pET151	CACCATGGCAACGGTGAGTGAAGT	TCACGATTCGTCGAAGTTCGTGC
P. fluor. Pf-5	PFL_3476	panE DHase	pET151	CACCATGAACATCACCATTCTTGGCG	CTACGAGGCGGCGCGTTTTTC
	PFL_3478	BFDC	pETSUMO	CGCGAACAGATTGGAGGTGGATCCAA AACCGTCCATAGCGCCT	GTGGCGGCCGCTCTATTAGGGTTC GATGGTCAGGGTC
	PFL_3481	Zn DHase	pET151	CACCATGCAAGCCATCAAGCTCAG	TCAGAGGCTCAGGCAGATTTTCCC

Assay of benzoylformate decarboxylase activity. 1. UV-visible assay. Initial enzymatic activities were measured using a coupled decarboxylase assay as described elsewhere (*19,20*). All the assays were performed at 25 ℃, in 50 mM potassium phosphate buffer pH 6.5 using a Cary 1 UV-vis spectrophotometer. The assay mixture was prepared in 1 ml 50 mM potassium phosphate buffer pH 6.5 containing benzoylformic acid (0.02-10 mM), NADH (3.5 mM), horse liver alcohol dehydrogenase (50 µg), thiamine diphosphate (0.5 mM), and $MgSO_4$ (2.5 mM). The reaction was started by addition of BFDC enzyme (2 µl, 0.3 mg/ml) and the decreasing absorption at 340 nm was examined. The specific extinction coefficient of NADH at 340 nm was 6.22 x 10^3 $M^{-1}cm^{-1}$. Kinetic parameters (k_{cat} and K_M) were determined by nonlinear curve fitting to the obtained enzyme activity, through Graphpad prism 8 software. In addition, data were fitted to the Hill equation to examine co-operativity.

2. HPLC assays. C-C cleavage and C-C bond formation reactions of BFDC were monitored by C_{18} reverse phase HPLC. Assays were performed in 50 mM potassium phosphate buffer pH 6.5, 25 ℃ in the presence of 5 mM substrate, 2.5 mM $MgSO_4$ and 0.5 mM thiamine diphosphate. Assays with recombinant glycolate oxidase and dehydrogenase enzymes were carried out in 50 mM potassium phosphate buffer pH 6.5, in the presence of 0.1 mM FMN (flavin mononucleotide) and 0.2 mM NAD^+, respectively. After reaction, 1 equivalent of MeOH was added to the enzyme assay to precipitate protein, and the mixture incubated for 10 min on ice, and then centrifuged (5 min, microcentrifuge) to remove protein, then

analysed by HPLC. The pH optimum of the decarboxylase reaction was measured in 1.5 mL batches containing 4-hydroxybenzoylformate (5 mM), $MgSO_4$ (2.5mM), ThDP (0.5mM), in 50 mM buffer at pH values between 4.0 and 9.0 using potassium acetate buffer (pH 4-5), potassium phosphate buffer (pH 6-8), and borate buffer (pH 8-9). The reaction was stopped after 1 h at 25°C.

HPLC analysis was conducted using a Phenomenex HyperClone™ 5 µm BDS C_{18} reverse phase column (130 Å, 5 µm, 4.6 mm) on a Hewlett-Packard Series 1100 analyser, at a flow rate of 0.8 mL/min, with monitoring at 270 or 254 nm. Method A (BFDC reactions): 0-5 min, 10% acetonitrile/water; 5-15 min, gradient 10-15% acetonitrile/water; 15-23 min, gradient 15-25% acetonitrile/water; 23-42 min, gradient 25-100% acetonitrile/water; 42-45 min, 100% acetonitrile; 45-50 min, gradient 100-10% acetonitrile/water. Method B (oxidase ro02984 reactions): 0-5 min, 15% MeOH/water; 5-15 min, gradient 15-25% MeOH/water; 15-23 min, gradient 25-70% MeOH/water; 23-30 min, gradient 70-100% MeOH/water; 30-35 min, 100% MeOH; 35-40 min, gradient 100-15% MeOH/water. 0.1% Trichloroacetic acid was added to solvents. Products were identified by comparison with authentic standards.

Hydrogen Peroxide (H_2O_2) assay for oxidase ro02984.

Amplex Red Hydrogen Peroxide/Peroxidase Assay kit (Invitrogen) was used to assay the hydrogen peroxide produced by oxidase (*21*). The assay was conducted according to the manufacturer's instruction, except the final concentrations of the H_2O_2 for standard curve were 0 to 2.5 µM. Fluorescence was detected at ex/em: 530/590 nm using a HIDEX sense microtiter plate reader. 0.5, 1, and 5 mM glycolate solutions with or without 0.1 mM FMN were prepared and incubated with Amplex® Red reagent/HRP working solution at room temperature protected from light. Fluorescence was immediately measured after 30 min.

Procedure for carboligation reaction catalysed by benzoylformate decarboxylase

Carboligation reactions were performed in 50 mM potassium phosphate buffer pH 6.5 at room temperature in the presence of 5 mM vanillin and 250 mM glycolaldehyde or acetaldehyde, 2.5 mM $MgSO_4$ and 0.5 mM thiamine diphosphate, to which were added RjBFDC or PfBFDC (0.15 mg), in a 2.0 mL total volume. HPLC analysis was carried out as described above.

2-Hydroxy-1-(4-hydroxy-3-methoxyphenyl)propan-1-one was identified as a product of BFDC-catalysed C-C bond formation reaction of acetaldehyde and vanillin at retention time 18 min, and was purified by C_{18} reverse phase HPLC. ^{1}H NMR (300 MHz, CD_3OD) δ_H 7.60 (1H, d, J = 7.0 Hz, aryl H-6), 7.56 (1H, s, aryl H-2), 6.85 (1H, d, J = 7.0 Hz, aryl H-5), 5.16 (1H, q, J = 6.5 Hz, $C_\beta\underline{H}(OH)$), 3.86 (3H, s, OCH_3), 1.40 (3H, d, $C_\gamma\underline{H_3}$) ppm. ^{1}H NMR spectrum shown in Supporting Information.

Crystallization, data collection, and structure determination of P. fluorescens BFDC

Pure recombinant PfBFDC (10 mg/mL) in 50 mM Tris pH 8, 150 mM NaCl, 0.1 mM TPP was subjected to crystallisation screening using a Mosquito liquid handling robot (TTP Labtech). 200 nL of protein was mixed with 200 nL of crystallisation solution from commercially available screens in MRC 96-well 2-drop crystallisation plates (Molecular Dimensions). Plates were sealed with sealing films (EXCEL Scientific) and incubated at 18 °C. Crystals appeared in a number of different conditions after 1 day. Rod shaped crystals were grown in condition E11 of the ShotGun (SG1) crystallisation screen (Molecular Dimensions) containing 0.1 M sodium acetate pH 4.6, 2M sodium formate. Crystals were removed from drops using a mounted Litholoop (Molecular Dimensions), cryoprotected in crystallisation solution containing 10% ethylene glycol, and flash-frozen in liquid nitrogen.

X-ray diffraction data to a resolution of 1.69 Å were collected at 100K at the beam line I04-1 at the Diamond Light Source, U.K. using a Pilatus 6M detector. All data were indexed, integrated and scaled using the XDS package (*22*). Further data handling was carried out using the CCP4 software package (*23*). The structure was solved by molecular replacement (starting from PDB:2FN3) using the automated pipeline by BALBES (*24*). Refinement of the structure was carried out by alternate cycles of manual refitting using Coot (*25*) and Refmac (*26*). Water molecules were added to the atomic model automatically using ARP (*27*) at the positions of large positive peaks in the difference electron density, only at places where the resulting water molecule fell into an appropriate hydrogen bonding environment. Restrained isotropic temperature factor refinements were carried out for each individual atom. The polypeptide chain was traced through electron density maps ($2mF_o$ -ΔF_c and mF_o – ΔF_c) for residues 2 to 526 (chain A) and 2 to 525 (chain B). Data collection and refinement statistics are given in **Table 2**.

Table 2. Crystallography data collection and refinement statistics

	PfBFDC
Data collection[a]	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c* (Å)	110.72, 130.72, 180.54
Wavelength (Å)	0. 91587
Resolution (Å)	85-1.69 (1.78-1.69)
Observations	1489650 (213925)
Unique reflections	144852 (10145)
R_{sym}	0.104 (1.097)
$I/\sigma(I)$	16.5 (2.0)
Completeness (%)	99.6 (97.5)
Redundancy	10.3 (10.4)
$CC_{1/2}$	0.999 (0.786)
Refinement	
R_{cryst}	0.159 (0.300)
Reflections used	139096 (9754)
R_{free}	0.180 (0.320)
Reflections used	5756 (391)
R_{cryst} (all data)	0.160
Non-hydrogen atoms	8696 (including 2 Na^+, 2 ThDP, 3 ethylene glycol and 796 water molecules
B-factors	
Protein	15.5
Water	28.4
R.m.s. deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.7
DPI coordinate error (Å)	0.079

[a]Values in parentheses are for highest-resolution shell.

Site directed mutagenesis. Site directed mutagenesis was performed for selected amino acids by using QuikChange II XL Site-Directed Mutagenesis Kit from Agilent technologies. For each target amino acid, a pair of primers containing the codon for the desired amino acid and 12 bp flanking sequence were designed, and the list of primers are shown below. The procedure was carried out according to manufacturer's instructions.

BFDC S73A: Fw: 5′- GAATCTGCACGCAGCTGCCGGCTCCGGCAATGCC -3′,

Rv: 5′- GGCATTGCCGGAGCCGGCAGCTGCGTGCAGATTC -3′

BFDC A73S: Fw: 5'-CAACCTGCACGCCGCGTCCGGCACCGGCAATGGC-3'

Rv: 5'-GCCATTGCCGGTGCCGGACGCGGCGTGCAGGTTG-3'

Results

Bioinformatic analysis of gene clusters in *R. jostii* RHA1 and *Pseudomonas fluorescens* Pf-5.

R. jostii RHA1 gene ro02986, found previously to encode vanillin dehydrogenase (*12*), is located within a cluster of 8 genes (ro02983 – ro02990) shown in Figure 1. Immediately downstream of ro02986 is gene ro02885 annotated as a benzoylformate decarboxylase, and gene ro02984 annotated as a flavin-dependent glycolate oxidase, and gene ro02983 encoding an αβ-hydrolase enzyme of unknown function. The cluster contains a further putative oxidase ro02988, two regulatory genes ro02987 and ro02990, and a putative MFS transporter ro02989.

Bioinformatic analysis of genomes of other potential bacterial lignin degraders revealed a similar gene cluster in *P. fluorescens* Pf-5, shown in Figure 1. This gene cluster contains a benzoylformate decarboxylase-like gene PFL_3478, an aldehyde dehydrogenase gene PFL_3475, and a *panE* dehydrogenase gene PFL_3476, normally found on the pantothenate biosynthetic pathway (*28*), and a second dehydrogenase gene PFL_3481. Nearby are *vanAB* genes PFL_3471 and PFL_3472 encoding vanillate demethylase, consistent with a pathway generating vanillin, which is then oxidised to vanillic acid (*12*), and also two MFS transporter genes PFL_3477 and PFL_3480 and a porin *oprD* gene PFL_3474. Smaller gene clusters containing benzoylformate decarboxylase and *panE* dehydrogenase genes were also found in the genomes of *Comamonas testosteroni* CNB-2 and *Burkholderia multivorans* ATCC 17616, as shown in Figure 1.

Amino acid sequence alignment between RjBFDC and PfBFDC proteins (Supporting Information Figure S1) showed that they shared 62% sequence identity, which suggested that they were structurally and functionally related. Gene RHA1_02986 encoding vanillin dehydrogenase shared 44% sequence identity with aldehyde dehydrogenases PFL_3475 in *P. fluorescens* and 42% sequence identity with BMULJ_02545 from *B. multivorans*.

Expression and purification of recombinant enzymes from *R. jostii* and *P. fluorescens*

Expression of recombinant *Rj*BFDC from vector pET 151 was optimised using 0.25 mM IPTG for induction of expression, protein aggregation being observed at higher concentrations of IPTG. A 1 L fermentation yielded 11.2 mg purified *Rj*BFDC, with the protein band for purified BFDC eluting at 55 kDa by SDS-PAGE (Supporting Information Figure S2), matching the predicted molecular weight (55.07 kDa). The recombinant enzyme was catalytically active for decarboxylation of benzoylformic acid, with k_{cat} 3.55 s^{-1} and K_M 0.49 mM (Supporting Information Figure S5). Purified enzyme only retained activity if purified in the presence of 0.1 mM thiamine diphosphate in the purification buffer.

The adjacent gene ro02984 in *R. jostii* RHA1 encoding glycolate oxidase was also amplified and was expressed optimally in pETite N-His SUMO Kan vector, since expression as a His_6 fusion protein was found to result in insoluble protein. A protein band was observed at the expected size (59 kDa) by SDS-PAGE in Fig. S2B (oxidase 47 kDa, Sumo tag 12 kDa), with a second protein band at 35 kDa also observed (Supporting Information Figure S2). oxidase ro02984 was expressed in the presence of 0.1 mM flavin mononucleotide, and the purified enzyme was yellow in colour (λ_{max} 375, 456 nm, see Supporting Information Figure S4), consistent with the presence of a flavin cofactor. The purified enzyme was active for glycolate oxidation, as determined by assay using Amplex Red. Activity was enhanced by addition of flavin mononucleotide, implying that the bound cofactor was FMN.

Three genes were cloned and expressed from *P. fluorescens* Pf-5, encoding dehydrogenase PanE, benzoylformate decarboxylase (PfBFDC) and Zn-dependent dehydrogenase PFL_3481. PanE and dehydrogenase PFL_3481 were expressed in vector pET151 and the recombinant proteins were expressed as His_6 fusion proteins, while PfBFDC was expressed in pETSUMO as a SUMO fusion protein. Recombinant PfBFDC was obtained in a yield of 8.4 mg purified enzyme from 1 L culture as a 55.9 kDa protein (see Supporting Information Fig S3), and the purified PfBFDC also showed ThDP-dependent activity toward benzoylformic acid with k_{cat} 12.1 s^{-1} and K_M 0.58 mM (Supporting Information Figure S5). Recombinant PfPanE and dehydrogenase PFL_3471 were each observed as the predicted 32 kDa and 34.7 kDa bands via SDS-PAGE (Supporting Information Figure S3). The pH-rate profiles for RjBFDC and PfBFDC showed optimum activity at pH 6.5 (see Supporting Information Figure S6).

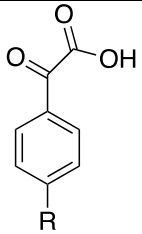
The steady-state kinetic data for the two benzoylformate decarboxylase enzymes were also fitted to the Hill equation to examine possible co-operative effects (see Supporting Information Figure S5). The Hill coefficient for RjBFDC was found to be 1.0, whereas the Hill coefficient for PfBFDC was 1.8, indicating that there may be co-operative binding in the PfBFDC enzyme.

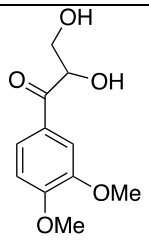
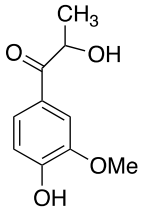
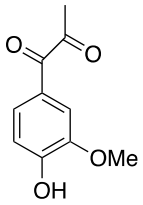
Investigation of hypotheses for enzyme function

A number of compounds related to hypotheses A and B were tested as substrates for RjBFDC and PfBFDC, using HPLC to examine enzyme-catalysed conversion, as shown in Table 3. Both enzymes were able to convert benzoylformic acid to benzaldehyde and carbon dioxide, using a UV-visible coupled assay, with specific activities of 41 and 78 U/mg for RjBFDC and PfBFDC respectively.

Both enzymes were also able to decarboxylate 4-hydroxybenzoylformic acid to 4-hydroxybenzaldehyde, monitored via HPLC assay. The specific activities for RjBFDC and PfBFDC were 12.3 and 24 U/mg at this pH, which were about 30% of the specific activities for benzoylformic acid (see Table 3).

Table 3 Specific activities (in μmol/min/mg protein) of BFDC and BFDC variants towards potential substrates under standard assay conditions. nd, not detected.

			RjBFDC	RjBFDC S73A	PfBFDC	PfBFDC A73S
	Aryl C2 substrate					
1	Benzoylformic acid (R = H)		41	48	78	7.6
2	4-Hydroxybenzoylformic acid (R = OH)		12.3	19	25	8.4
	Aryl C3 substrate					

3	C-Veratroylglycol		nd.	nd.	nd.	nd.
4	2-hydroxy-1-(4-hydroxy-3-methoxyphenyl)propan-1-one		nd.	nd.	nd.	nd.
5	1-(4-hydroxy-3-methoxyphenyl)propane-1,2-dione		nd.	nd.	nd.	nd.

Hypothesis A was tested using aryl C_3 keto-diol C-veratroylglycol (**3**) and diketone 1-(4-hydroxy-3-methoxyphenyl)propane-1,2-dione (**5**). No conversion of either compound by RjBFDC or PfBFDC could be detected by HPLC analysis. The reverse carboligase reactions catalysed by BFDC were investigated using vanillin and C_2 aldehydes glycolaldehyde and acetaldehyde. Using glycoaldehyde, no formation of keto-diol **3** was detected even after 24 hr, however, using 5 mM vanillin and 250 mM acetaldehyde as substrates, RjBFDC was able to catalyse the formation of ketol **4** (Figure 3), whose structure was confirmed by ^{1}H NMR spectroscopy (Supporting Information Figure S7). The specific activity for the carboligation reaction catalysed by RjBFDC was 0.27 μmol/min/mg protein, compared with specific activities of 41 and 12.3 μmol/min/mg protein for decarboxylation of benzoylformate and 4-hydroxybenzoylformate respectively. PfBFDC also catalysed the carboligation reaction, with a higher specific activity of 0.75 μmol/min/mg protein.

Recombinant *P. fluorescens* dehydrogenases PanE and PFL_3481 and *R. jostii* oxidase ro02984 were incubated with several aromatic hydroxyl carboxylic acids, and product formation monitored via reverse phase HPLC analysis. An authentic sample of guaiacyl-glycerol was synthesised (see hypothesis A), using the method of Adler and Yllner (*17*), and its structure verified by ^{1}H NMR spectroscopy. However, incubation with Pfl dehydrogenases PanE or PFL_3481 and NAD^+ or $NADP^+$ showed no activity for

oxidation of guaiacyl-glycerol, and neither did *R. jostii* oxidase ro02984 show activity for this compound. Furthermore, guaiacyl-glycerol was unable to support growth of *R. jostii* RHA1 as a carbon source in M9 minimal media, at 0.1% (w/w) concentration, which might be expected if it was an intermediate in hypothesis A. Hence our data suggest that neither guaiacyl-glycerol nor the oxidised keto-diol in Hypothesis A appear to be intermediates in the production of vanillin in *R. jostii*.

Hypothesis B involves the oxidation of either a substituted mandelic acid or a substituted phenylglyoxal to 3-methoxy-4-hydroxybenzoylformic acid, catalysed by oxidase ro02984 (see Figure 2). Recombinant oxidase ro02984 was able to convert mandelic acid into benzoylformic acid, with specific activity 15 μmol/min/mg protein (Supporting Information Figure S9), however, it showed no activity for oxidation of 4-hydroxymandelic acid, nor for 3-methoxy-4-hydroxymandelic acid. Pfl dehydrogenases PanE and PFL_3481 also showed no activity for oxidation of 4-hydroxymandelic acid. Phenylglyoxal was also oxidised by oxidase ro02984, to give benzoylformic acid as a major product (specific activity 32 μmol/min/mg protein), and benzoic acid as a minor product (specific activity 0.23 μmol/min/mg, see Supporting Information Figure S9). Pfl dehydrogenases PanE and PFL_3481 also showed no activity for oxidation of phenylglyoxal.

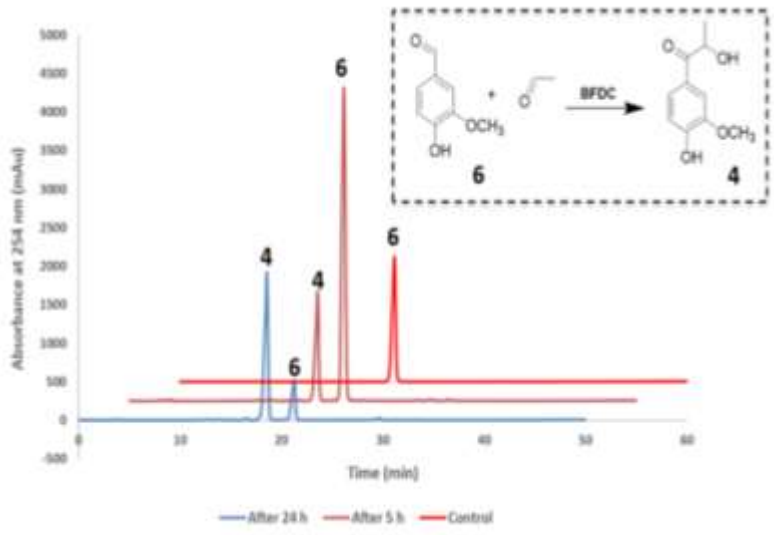


Figure 3. Reverse phase HPLC analysis of *P.f*BFDC incubation with 5 mM vanillin and 250 mM acetaldehyde after 5 h (brown) and 24 h (blue), showing the formation of a new product peak at 18 min. Control (red): without enzyme.

Crystal Structure of PfBFDC

The crystal structure of *Pseudomonas fluorescens* benzoylformate decarboxylase (PfBFDC) has been determined at 1.69 Å resolution (PDB:6QSI). The asymmetric unit of the crystal contained two molecules with the space group $C222_1$. The structure shows PfBFDC to be a tetramer, more specifically a dimer of active dimers, which is generated by crystallographic symmetry (Fig. 4a). Each individual subunit is organised into three separate domains, as has been observed for previously published benzoylformate decarboxylase (BFDC) structures 1BFD (*29*) and 4Q9D (*30*) and other thiamine diphosphate (ThDP) – dependent enzymes pyruvate decarboxylase 1PVD (*31*), oxalyl CoA decarboxylase 2VBF (*32*), and indolepyruvate decarboxylase 1OVM (*33*). The three domains of PfBFDC are as follows; domain one (residues 2-182), domain two (residues 183-343) and domain three (residues 344-526) (Fig. 4c).

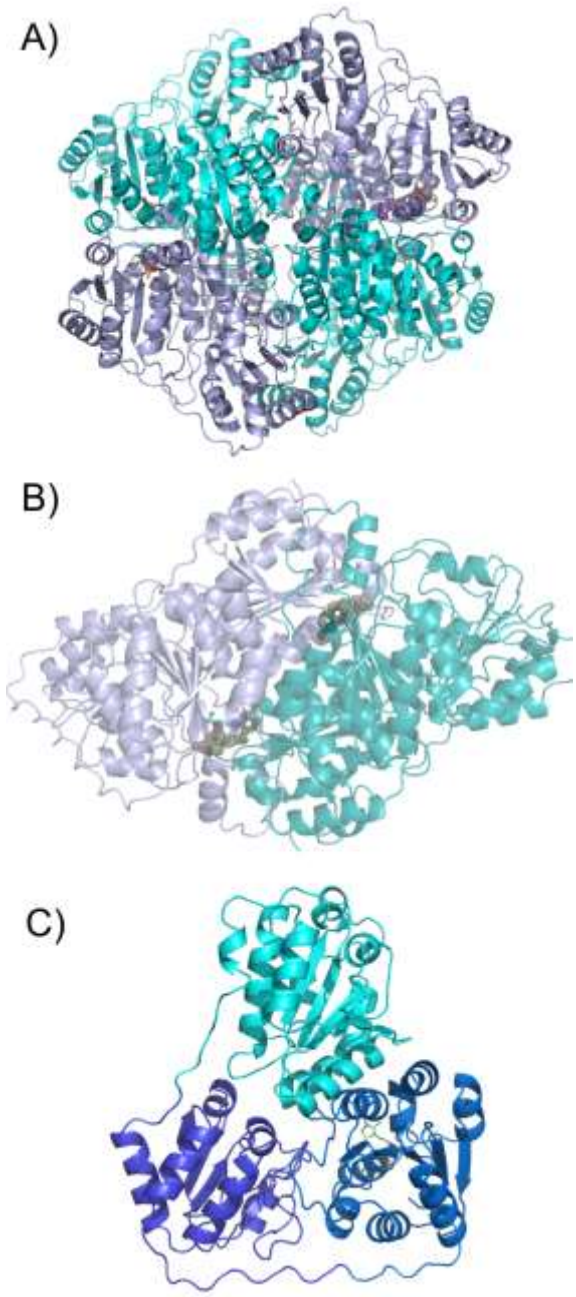


Figure 4: Structure of PfBFDC. (A) Dimer of active dimers for PfBFDC, as shown using symmetry mates. (B) PfBFDC dimer with the two ThDP molecules with their corresponding density ($2mF_o$ -ΔF_c map at σ level 2) shown at the interface of the monomers. (C) Domain layout for PfBFDC monomer: domain 1 (cyan), domain 2 (bright blue) and domain 3 (deep blue). Protein is depicted as cartoon with ThDP shown as sticks.

PfBFDC Active Site

Each active dimer contains two molecules of thiamine diphosphate (ThDP) in the V-conformation previously noted in other ThDP-dependent enzymes (*29-35*), at the

interface between the two subunits (Fig. 4b). The ThDP molecule itself is situated within domain 3 of one of the subunits (Fig. 4c) but is co-ordinated between domain 3 of one subunit and domain 1 from the other subunit within the active dimer. The thiamine group is directly co-ordinated at N1' by Glu47 and Gln23 from one subunit, as described previously for *P. putida* BFDC (PpBFDC) (*29*). The main chain carbonyl of Gly401 from the other subunit within the dimer interacts with the N4' of the thiamine group (Fig. 5a). The phosphate moieties of ThDP are co-ordinated by the hydroxyl groups of Ser430, Ser378 and Thr377. The co-ordination of a water molecule between His70 and Ser26 is observed within the active site (Fig. 5b). A water molecule at this position has previously been shown to be involved in catalysis in PpBFDC (*29*) with mutants H70A and S26A showing a reduction in k_{cat} and increased K_M, respectively, compared to wild type enzyme (*36*).

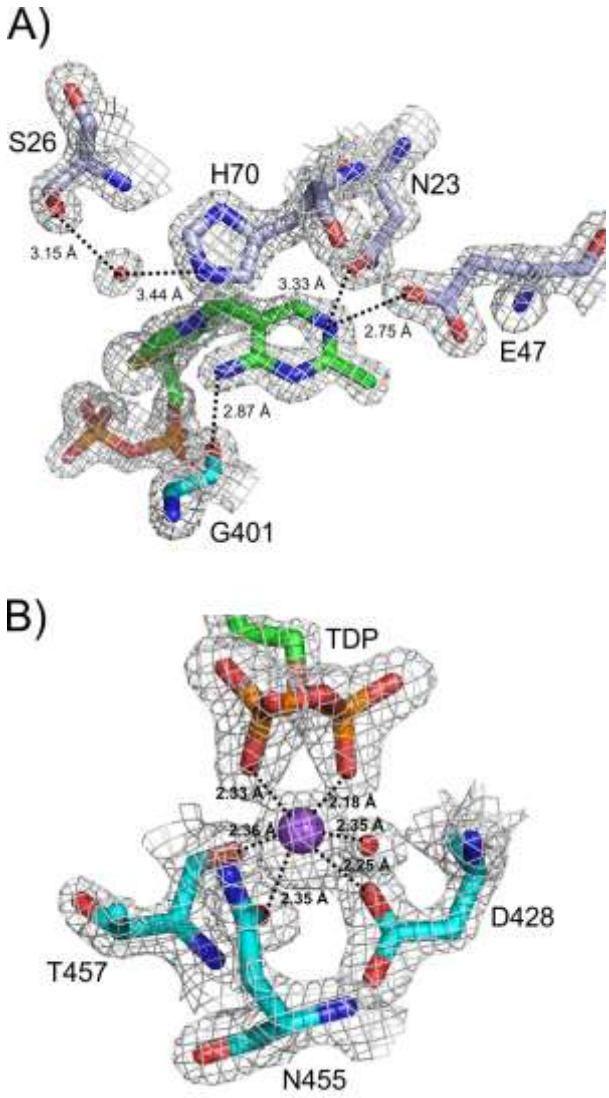


Figure 5: Co-ordination of ThDP and sodium ion in the active site of PfBFDC. (A) Interactions between key active site residues and ThDP and the active site water molecule

and density ($2mF_o$ -ΔF_c map at σ level 2.0). (B) Octahedral co-ordination of sodium ion and density ($2mF_o$ -ΔF_c map at σ level 2.0). Residues and ThDP are shown as sticks (cyan for subunit 1 and light blue for subunit 2), water (red) and sodium (purple) ions are shown as spheres.

As with other ThDP-dependent enzymes, observed density supported the octahedral co-ordination of a metal ion at the ThDP binding site (*29,31,33*). Analysis with CheckMyMetal server (*37*) suggests that the metal ion is a sodium ion (see Supporting Information Figure S11), which was the only metal ion used in the crystallisation buffer, suggesting that this is likely an artefact of crystallisation. Previously, ThDP-dependent enzymes have been shown to contain Ca^{2+} (*29*) or Mg^{2+} (*31,33*) in the active site. The sodium ion is co-ordinated by two phosphate oxygens from the ThDP cofactor, the hydroxyl groups of Asp428 and Asn455, the main chain carbonyl of Thr457, and a water molecule (Fig. 5b), similar to the metal co-ordination observed in other ThDP-dependent enzymes (*29,33*).

Comparison with other ThDP-Dependent Enzymes

Sequence alignment of PfBFDC with other ThDP-dependent protein structures (PDB: 1BFD, 4K9Q, 4Q9D, 2JI6, 2Q5J, 1PVD, 2VBF, 1OVM, and 2VBI) shows that PfBFDC clusters alongside the other benzoylformate decarboxylases away from other ThDP-dependent enzymes. Structural alignments using SuperPose (*38*) of PfBFDC with other solved BFDC structures, gives a global root-mean-square deviation (RMSD) over alpha carbons of 0.35 Å for 1BFD (*P. putida* BFDC), 0.65 Å for 4K9Q (*P. necessarius* BFDC), and 0.78 Å or 4Q9D (*M. smegmatis* BFDC). These structural comparisons suggest that the enzymes fold in a highly similar manner despite slight variations in the individual protein's amino acid sequence.

Superposition of PpBFDC (PDB: 1BFD) with PfBFDC highlights the similarity across the main three domains of the monomer, with major differences in flexible loops connecting the domains. Previously identified active site residues for the decarboxylation mechanism in PpBFDC are Ser-26, His-70 and His-281 (*36,39,40*). His-70 and His-281 were initially predicted to act as acid-base catalytic residues, and Ser26 involved in substrate binding (*36*). Further mutational studies showed that the effect of replacement of His-281 or His-70 on catalysis is dependent on the amino acid substituted (*40*). Further

work using density functional theory calculations highlighted the probability that His-70, 281 and Ser-26 are all involved in substrate binding via hydrogen bonding (*41*). Superposition of PpBFDC and PfBFDC active sites shows the position of most residues identified as important in catalysis in PpBFDC overlay with near identical positioning to those in PfBFDC (Fig. 6). Interestingly, the orientation of His-281 differs between these two active sites (Fig. 6). His-281 has been shown to be important to PpBFDC catalysis and substrate binding (*29*,*36*,*40*), therefore this alteration in orientation could be important in differential activity or substrate specificity between the enzymes.

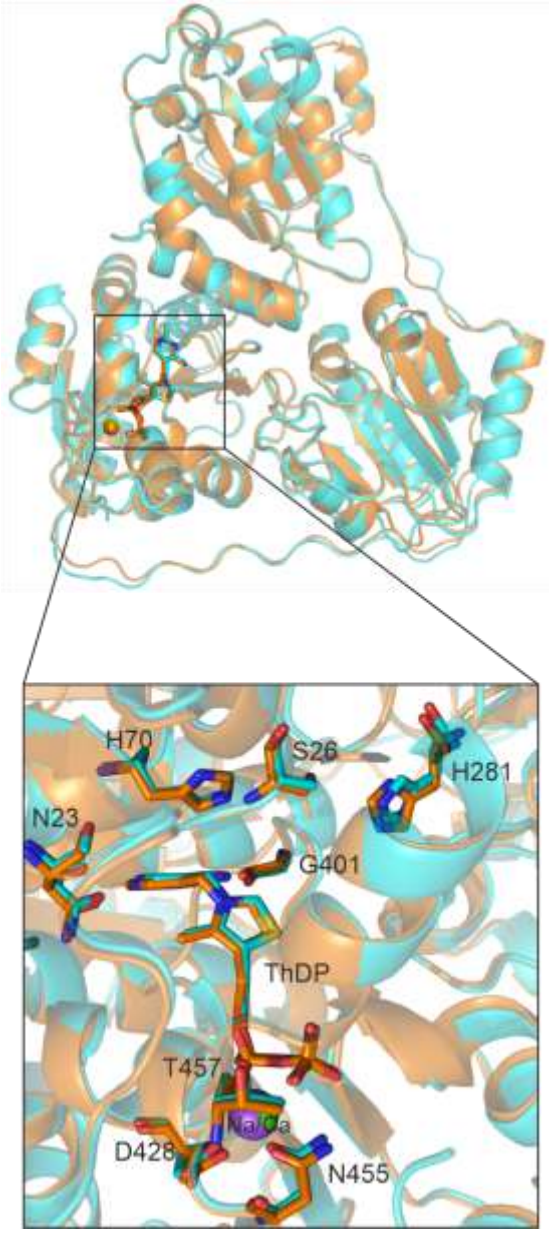


Figure 6: Superposition of *P. putida* BFDC (1BFD, chain A) in orange and *P. fluorescens* BFDC (PfBFDC) in cyan. Protein is shown as cartoon and key residues in ThDP co-ordination and ThDP shown as sticks, and metal ions are shown as spheres.

Investigation of role of residue 73 in C-C bond formation activity

RjBFDC and PfBFDC showed different carboligase activity towards vanillin and acetaldehyde. As shown in Figure 7, RjBFDC catalysed a 30% conversion of vanillin and acetaldehyde to ketol product **4** after 10 hr, however, PfBFDC catalysed 70% conversion under the same conditions. Benzoylformate decarboxylase from *Pseudomonas putida* has been reported as an effective catalyst for carboligation reactions with benzaldehyde substrates, although not with vanillin (*42*), so it was of interest to explore the difference in reactivity between these two enzymes.

Comparison of the crystallographic evidence and sequence alignment of both enzymes identified one active site residue that was different between the two enzymes: Ala-73 in the PfBFDC structure (situated 9.1 Å from C-2 of the ThDP cofactor, see Supporting Information Figure S10) is replaced by Ser-73 in RjBFDC. Ala-73 is also found in *P. putida* benzoylformate decarboxylase (*33*).

Site-directed mutagenesis was therefore carried out to replace this residue in both RjBFDC and PfBFDC: Ser-73 in RjBFDC was replaced with alanine, and Ala-73 in PfBFDC was replaced with serine. The mutants S73A RjBFDC and A73S PfBFDC were overexpressed in *E. coli*, and the recombinant proteins were purified using the methodology described for the wild-type enzyme. As described in Table 3, compared to wild-type enzymes, the specific activities of S73A RjBFDC toward benzoylformic acid and 4-hydroxybenzoylformic acid increased, while A73S PfBFDC lost >60% activity. Carboligation activity of the mutant enzymes with vanillin and acetaldehyde as substrates was also examined by HPLC analysis, as shown in Figure 7. The specific activity with variant containing the S73A RjBFDC mutation increased 2-3 fold, and A73S PfBFDC mutation was found to lose nearly all its carboligase activity (see Figure 7).

These findings demonstrate that the residue at position 73 influences carboligation activity, with a marked preference for Ala vs Ser in both enzymes. Replacement of Ala with Ser narrows the space within the binding site of both the donor and acceptor aldehyde, hence steric factors may be predominantly responsible for the differences in activity.

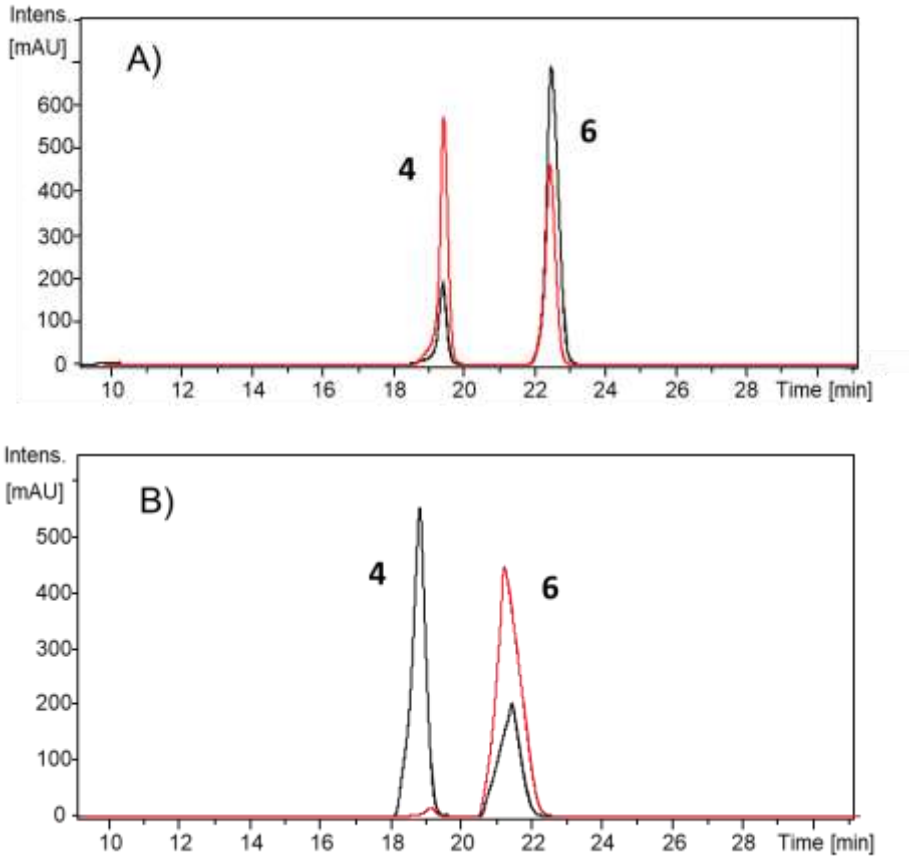


Figure 7. Reverse phase HPLC analysis of A) WT RjBFDC (gray) and S73A mutant (red) B) WT PfBFDC (gray) and A73S mutant (red) incubated with 5 mM vanillin and 250 mM acetaldehyde after 10 h. **4**: 2-hydroxy-1-(4-hydroxy-3-methoxyphenyl)propan-1-one, **6**: vanillin.

Discussion

The *R. jostii* gene cluster containing ro02985, annotated as a benzoylformate decarboxylase gene, is implicated in lignin degradation, because gene knockout of the vanillin dehydrogenase gene ro02986 gives a mutant strain that accumulates vanillin when grown on minimal media containing wheat straw lignocellulose. The gene cluster containing ro02985 is different to that containing benzoylformate decarboxylase in *Pseudomonas putida*, which also contains mandelate racemase and mandelate dehydrogenase genes for mandelic acid utilisation (*14*). In this study we have examined two hypotheses for the function of a thiamine diphosphate-dependent enzyme in this cluster, also found in a related cluster in *Pseudomonas fluorescens* Pf-5.

Hypothesis A involves a transketolase-type cleavage of a keto-diol aryl C_3 intermediate **3**, to generate vanillin and glycoaldehyde, however, neither RjBFDC nor

PfBFDC were able to process an authentic sample of **3**. Furthermore, the proposed intermediate guaiacyl-glycerol was not accepted as a substrate by Rj glycolate oxidase, nor by the Pfl dehydrogenases PanE or PFL_3481, nor could this compound support growth of *R. jostii* RHA1 on minimal media.

The alternative hypothesis B is consistent with the decarboxylation of benzoylformate and 4-hydroxybenzoylformate by RjBFDC and PfBFDC. 4-Hydroxylbenzoylformate could potentially be formed via oxidation of either a substituted mandelic acid, or a 2-keto-aldehyde precursor. Neither oxidase ro02984 nor the two Pfl dehydrogenases could oxidise 4-hydroxymandelic acid, although oxidase ro02984 can oxidise unsubstituted mandelic acid. Oxidase ro02984 was found to oxidise phenylglyoxal to benzoylformic acid and benzoic acid, suggesting that it could oxidise a substituted phenylglyoxal as proposed in Hypothesis B. Oxidative decarboxylation of a benzoylformic acid to a benzoic acid by flavin is precedented by the work of Shinkai *et al* (*43*). Therefore, the observed enzyme activities are consistent with Hypothesis B, and we propose that 4-hydroxybenzoylformate is formed during lignin degradation via oxidation of a keto-aldehyde precursor. The activity of oxidase ro02984 for mandelic acid oxidation suggests that this gene cluster might also be used for mandelic acid breakdown in *R. jostii* RHA1, but not for substituted mandelic acids. We have as yet not found an enzymatic activity for Pfl dehydrogenases PanE or PFL_3481.

The formation of an aryl C_2 keto-aldehyde precursor, proposed in Hypothesis B, could occur via oxidative cleavage of phenylcoumaran or diarylpropane units found in polymeric lignin (see Figure 2A for structures), via the pathway proposed in Figure 8. Benzylic oxidation of phenylcoumaran units is plausible, and has been observed by Shettigar et al. in a *Pseudomonas* strain that can degrade pinoresinol, another component of lignin (*44*). Oxidation of the γ-hydroxyl group, followed by decarboxylation, would give a ketol intermediate. Aryl-Cα cleavage of such an intermediate, as observed with a lignin-oxidising manganese superoxide dismutase from *Sphingobacterium* sp. T2 (*18*), would generate the proposed keto-aldehyde precursor.

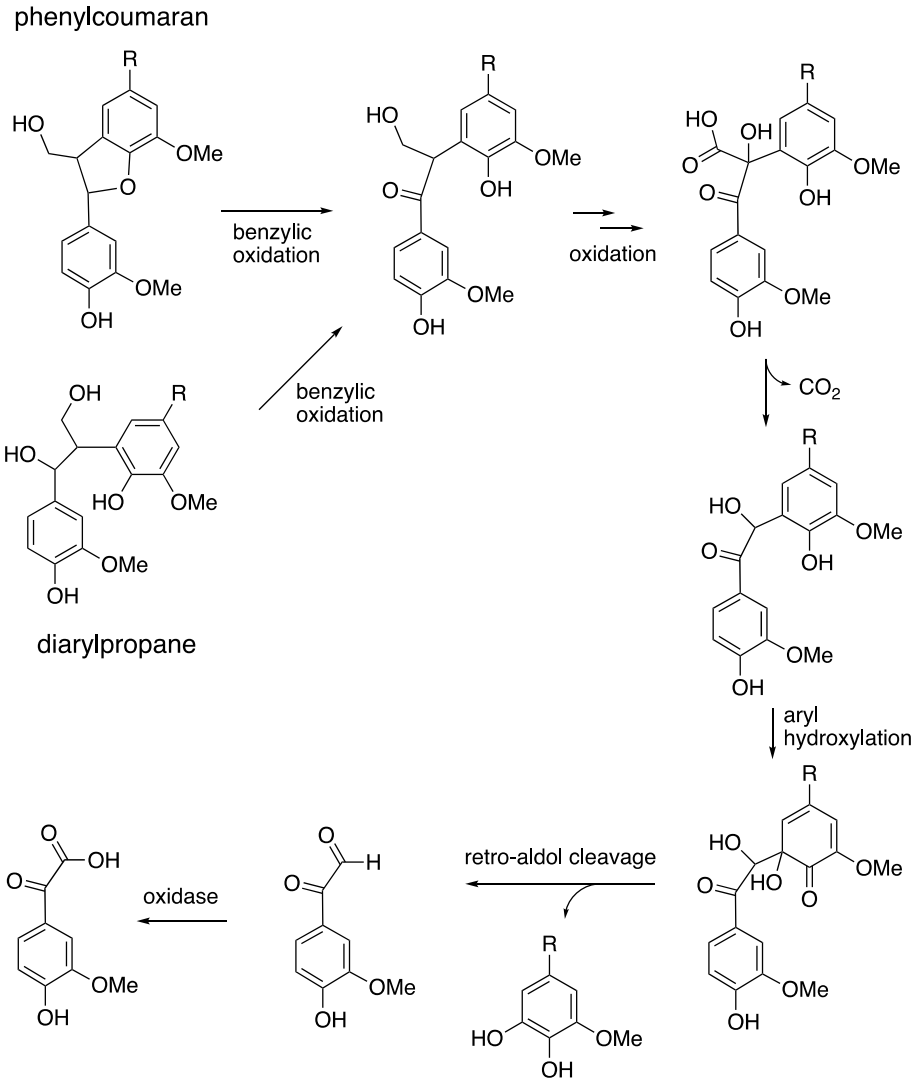


Figure 8. Possible pathway for conversion of phenylcoumaran and diarylpropane units in lignin into an aryl C_2 keto-aldehyde intermediate. R, lignin polymer.

Determination of the crystal structure of PfBFDC shows that these enzymes possess the same catalytically important amino acids studied in BFDC enzymes from other bacteria. Examination of the X-ray structure has identified position 73 in BFDC as a difference between RjBFDC and PfBFDC active sites, and we have shown via site-directed mutagenesis that the presence of alanine vs serine is important for the carboligation activity of RjBFDC. Potentially this observation could be rationalised by the smaller alanine sidechain allowing binding of an acetaldehyde co-substrate, but not the slightly larger glycoaldehyde. We note that this residue lies outside the binding pockets identified in *P. putida* BFDC by Gocke et al (*45*), suggesting either that a different binding mode is used in RjBFDC, or that this mutation causes a subtle structural change.

The present study provides evidence for the first time for a pathway involving metabolism of aryl C_2 lignin degradation fragments that is used by lignin-degrading bacteria. This pathway also detoxifies aldehyde products, and generates hydrogen peroxide from oxidase activity, which could then be used for lignin oxidation by DyP-type peroxidases (*4*-*6*). In basidiomycete fungi, oxidation of aldehyde lignin fragments occurs via a copper-dependent glycol oxidase enzyme that is thought to act as an accessory enzyme for lignin degradation (*46*), so the oxidase enzyme identified here may fulfil a similar role.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement 705423 and grant agreement 720303, and by BBSRC research grants BB/M025772/1 and BB/M003523/1. Crystallographic data were collected at beamline I04-1 at Diamond Light Source, UK and we acknowledge the support of the beamline scientist Dr Frank von Delft. We thank Shanshan Zhou (University of Warwick) for assistance with chemical synthesis work, Dr Rahman Rahmanpour for assistance with molecular genetics work, and Dr. Lan Wu and Prof. Jeremy Luterbacher (EPFL) for the gift of a sample of 1-(4-hydroxy-3-methoxyphenyl)propane-1,2-dione.

Supporting Information

Supporting information is available, which comprises: amino acid sequence alignment for BFDC enzymes (Figure S1); SDS-PAGE gels for recombinant enzymes (Figures S2, S3); UV/vis spectrum for recombinant *R. jostii* oxidase ro02984 (Figure S4); steady-state kinetic data for BFDC enzymes (Figures S5, S6); NMR spectra for 2-hydroxy-1-(4-hydroxy-3-methoxyphenyl)propan-1-one and guaiacylglycerol (Figures S7, S8); HPLC data for substrate conversions by oxidase ro02984 (Figure S9); superposition of PfBFDC and RjBFDC structures (Figure S10); and metal:ligand co-ordination values for PfBFDC structure (Figure S11).

References

1. Bugg, T. D. H. and Rahmanpour, R. (2015) Enzymatic conversion of lignin into renewable chemicals, *Curr. Opin. Chem. Biol. 29*, 10-17.

2. Ten Have R. and Teunissen P.J.M. (2001) Oxidative mechanisms involved in lignin degradation by white-rot fungi. *Chem. Rev*. *101*, 3397-3413.
3. Wong D.W.S. (2009) Structure and action mechanism of lignolytic enzymes. *Appl. Biochem. Biotechnol 157*, 174-209.
4. Ahmad, M., Roberts, J. N., Hardiman, E. M., Singh, R., Eltis, L. D. and Bugg, T. D. H. (2011) Identification of DypB from *Rhodococcus jostii* RHA1 as a lignin peroxidase, *Biochemistry, 50*, 5096-5107.
5. Brown, M. E., Barros, T. and Chang, M. C. (2012) Identification and characterization of a multifunctional dye peroxidase from a lignin-reactive bacterium, *ACS Chem. Biol. 7*, 2074-2081.
6. Rahmanpour, R. and Bugg, T. D. H. (2015) Characterisation of Dyp-type peroxidases from *Pseudomonas fluorescens* Pf-5: Oxidation of Mn(II) and polymeric lignin by Dyp1B, *Arch. Biochem. Biophys. 574*, 93-98.
7. Majumdar, S., Lukk, T., Solbiati, J. O., Bauer, S., Nair, S. K., Cronan, J. E. and Gerlt, J. A. (2014) Roles of small laccases from *Streptomyces* in lignin degradation, *Biochemistry 53*, 4047-4058.
8. Granja-Travez, R.S., Wilkinson, R.C., Persinoti, G.F., Squina, F.M., Fülöp, V., and Bugg, T.D.H. (2018) Structural and functional characterisation of a multi-copper oxidase CueO from lignin-degrading bacterium *Ochrobactrum* sp. reveal its activity towards lignin model compounds and lignosulfonate, *FEBS J. 285*, 1684-1700.
9. Jimenez, J.I., Minambres, B., Garcia, J.L., and Diaz, E. (2002). Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environ. Microbiol. 4*, 824-841.
10. Bugg T.D.H., Ahmad M., Hardiman E.M., and Rahmanpour R. (2011) Pathways for degradation of lignin in bacteria and fungi. *Nat. Prod. Reports 28*, 1883-1896.
11. Bugg, T.D.H., Ahmad, M., Hardiman, E.M., and Singh, R. (2011) The emerging role for bacteria in lignin degradation and bio-product formation. *Curr. Opin. Biotechnol. 22*, 394-400.
12. Chen, H-P., Chow, M., Liu, C.C., Lau, A., Jiu, J., and Eltis, L.D. (2012) Vanillin catabolism in *Rhodococcus jostii* RHA1. *Appl. Environ. Microbiol. 78*, 586-588.
13. Sainsbury, P. D., Hardiman, E. M., Ahmad, M., Otani, H., Seghezzi, N., Eltis, L. D. & Bugg, T. D. (2013) Breaking down lignin to high-value chemicals: the conversion of

lignocellulose to vanillin in a gene deletion mutant of *Rhodococcus jostii* RHA1, *ACS Chem. Biol. 8*, 2151-2156.

14. Tsou, A.Y., Ransom, S.C., Gerlt, J.A., Buechter, D.D., Babbitt, P.C., and Kenyon, G.L. (1990) Mandelate pathway of *Pseudomonas putida*: sequence relationships involving mandelate racemase, (S)-mandelate dehydrogenase, and benzoylformate decarboxylase and expression of benzoylformate decarboxylase in *Escherichia coli***.** *Biochemistry 29*, 9856–9862.

15. Taylor, C. R., Hardiman, E. M., Ahmad, M., Sainsbury, P. D., Norris, P. R. and Bugg, T. D. H. (2012) Isolation of bacterial strains able to metabolize lignin from screening of environmental samples, *J. Appl. Microbiol. 113*, 521-530.

16. Turner, N.J. (2000) Applications of transketolases in organic synthesis. *Curr. Opin. Biotechnol. 11,* 527-531.

17. Adler, E. and Yllner, S. (1953) Synthesis and reactions of α-(3-methoxy-4-hydroxyphenyl)-glycerol (guaiacyl-glycerol). II. Synthesis. *Acta Chem. Scand*. *7*, 570-581.

18. Rashid, G. M. M., Taylor, C. R., Liu, Y., Zhang, X., Rea, D., Fülöp, V., and Bugg, T. D. H. (2015) Identification of manganese superoxide dismutase from *Sphingobacterium* sp. T2 as a novel bacterial enzyme for lignin oxidation. *ACS Chem. Biol. 10*, 2286-2294.

19. Weiss P. M., Garcia G. A., Kenyon G. L., Cleland W. W., and Cook P. F. (1988) Kinetics and mechanism of benzoylformate decarboxylase using ^{13}C and solvent deuterium isotope effects on benzoylformate and benzoylformate analogues. *Biochemistry 27*, 2197-2205.

20. Pohl M., Grötzinger J., Wollmer A., and Kula M. R. (1994) Reversible dissociation and unfolding of pyruvate decarboxylase from *Zymomonas mobilis*. *Eur. J. Biochem*. *224*, 651-661.

21. Votyakova T.V. and Reynolds I.J. (2004) Detection of hydrogen peroxide with Amplex Red: interference by NADH and reduced glutathione auto-oxidation. *Arch. Biochem. Biophys. 431*, 138-144.

22. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D* **66**, 125-132

23. Dodson, E. J., Winn, M., and Ralph, A. (1997) [32] Collaborative computational project, number 4: Providing programs for protein crystallography. *Methods in Enzymology 277*, 620-633

24. Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallographica Section D* **64**, 125-132

25. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D 60*, 2126-2132
26. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr. 53*, 240-255
27. Langer, G. G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols 3*, 1171-1179
28. Zheng, R. and Blanchard, J.S. (2000) Kinetic and mechanistic analysis of the *E. coli panE*-encoded ketopantoate reductase. *Biochemistry 39*, 3708-3717.
29. Hasson M. S., Muscate A., McLeish M. J., Polovnikova L. S., Gerlt J. A., Kenyon G. L., Petsko G. A. and Ringe D. (1998) The crystal structure of benzoylformate decarboxylase at 1.6 Å resolution: diversity of catalytic residues in thiamin diphosphate-dependent enzymes. *Biochemistry 37*, 9918-9930
30. Andrews F. H., Horton J. D., Shin D., Yoon H., Logsdon M. G., Malik A. M., Rogers M. P., Kneen M. M., Suh, S.W., and McLeish M. J. (2015) The kinetic characterization and X-ray structure of a putative benzoylformate decarboxylase from *M. smegmatis* highlights the difficulties in the functional annotation of ThDP-dependent enzymes. *Biochim. Biophys. Acta 1854*, 1001-1009.
31. Arjunan P., Umland T., Dyda F., Swaminathan S., Furey W., Sax M., Farrenkopf B., Gao Y., Zhang D., and Jordan F. (1996) Crystal structure of the thiamin diphosphate-dependent enzyme pyruvate decarboxylase from the yeast *Saccharomyces cerevisiae* at 2.3 Å resolution. *J. Mol. Biol. 256*, 590-600.
32. Berthold C. L., Toyota C. G., Moussatche P., Leeper F.J., Richards N. G. J. and Lindqvist Y. (2007) Crystallographic snapshots of oxalyl-CoA decarboxylase give insights into catalysis by nonoxidative ThDP-dependent decarboxylases. *Structure 15*, 853 – 861.
33. Shütz, A., Sandalova, T., Ricagno, S., Hübner, G., König, S., and Schneider, G. (2003) Crystal structure of thiamindiphosphate-dependent indolepyruvate decarboxylase from Enterobacter cloacae, an enzyme involved in the biosynthesis of the plant hormone indole-3-acetic acid. *Eur. J. Biochem. 270*, 2312-2321.
34. Versées W., Spaepen S., Wood M. D. H., Leeper F. J., Vanderleyden J., and Steyaert J. (2007) Molecular mechanism of allosteric substrate activation in a thiamine diphosphate-dependent decarboxylase. *J. Biol. Chem. 282*, 35269-35278

35. Berthold C. L., Gocke D., Wood M. D., Leeper F. J., Pohl M. and Schneider G. (2007) Structure of the branched-chain keto acid decarboxylase (KdcA) from *Lactococcus lactis* provides insights into the structural basis for the chemoselective and enantioselective carboligation reaction. *Acta crystallographica. Section D 63*, 1217-1224.
36. Polovnikova E. S., McLeish M. J., Sergienko E. A., Burgner J. T., Anderson N. L., Bera A. K., Jordan F., Kenyon G. L. and Hasson M. S. (2003) Structural and kinetic analysis of catalysis by a thiamin diphosphate-dependent enzyme, benzoylformate decarboxylase. *Biochemistry 42*, 1820-1830.
37. Zheng H., Cooper D. R., Porebski P. J., Shabalin I. G., Handing K. B. and Minor W. (2017) CheckMyMetal: a macromolecular metal-binding validation tool. *Acta crystallographica. Section D, Structural biology 73*, 223-233.
38. Maiti R., Van Domselaar G.H., Zhang H., and Wishart D.S. (2004) SuperPose: a simple server for sophisticated structural superposition. *Nucl. Acids Res. 32*, W590-W594.
39. Bera A. K., Polovnikova L. S., Roestamadji J., Widlanski T. S., Kenyon G. L., McLeish M. J., and Hasson M. S. (2007) Mechanism-based inactivation of benzoylformate decarboxylase, a thiamine diphosphate-dependent enzyme*. J. Am. Chem. Soc. 129*, 4120-4121
40. Yep A., Kenyon G. L. and McLeish M. J. (2008) Saturation mutagenesis of putative catalytic residues of benzoylformate decarboxylase provides a challenge to the accepted mechanism. *Proc. Natl. Acad. Sci. USA* 105, 5733-5738.
41. Planas F., Sheng X., McLeish M. J. and Himo, F. (2018) A theoretical study of the benzoylformate decarboxylase reaction mechanism. *Frontiers in Chemistry 6*, article 205.
42. Iding, H., Dünnwald, T., Greiner, L., Liese, A., Müller, M., Siegert, P., Grötzinger, J., Demir, A.S., and Pohl, M. (2000) Benzoylformate decarboxylase from *Pseudomonas putida* as a stable catalyst for the synthesis of chiral 2-hydroy ketones. *Chem. Eur. J. 6*, 1483-1495.
43. Shinkai, S., Yamashita, T., and Manabe, O. (1979) Flavin-mediated oxidative decarboxylation of p-chlorobenzoylformic acid assisted by cyanide ion and cationic micelles. *J. Chem. Soc. Chem. Commun*., 301-302.
44. Shettigar, M., Balotra, S., Cahill, D., Warden, A.C., Lacey, M.J., Kohler, H-P.E., Rentsch, D., Oakeshott, J.G., and Pandey, G. (2018) Isolation of the (+)-pinoresinol-mineralizing *Pseudomonas* sp. strain SG-MS2 and elucidation of its catabolic pathway. *Appl. Environ. Microbiol*. *84*, e02531-17.

45. Gocke, D., Walter, L., Gauchenova, E., Kolter, G., Knoll, M., Berthold, C.L., Schneider, G., Pleiss, J., Müller, M., and Pohl, M. (2008) Rational design of ThDP-dependent enzymes – engineering stereoselectivity. *ChemBioChem 9*, 406-412.
46. Daou, M and Faulds, C.B. (2017) Glyoxal oxidases: their nature and properties. *World J. Microbiol. Biotechnol*. *33*, 87.

Accession codes

Enzyme	Gene ID	Uniprot
R. jostii benzoylformate decarboxylase	RHA1_ro02985	Q0SCE8
R. jostii oxidase ro02984	RHA1_ro02984	Q0SCE9
P. fluorescens benzoylformate decarboxylase	PFL_3478	Q4KB02
P. fluorescens PanE dehydrogenase	PFL_3476	Q4KB04
P. fluorescens dehydrogenase PFL_3481	PFL_3481	Q4KAZ9